ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/344487215

A Review On Sentiment Analysis Methodologies, Practices And Applications

Research · October 2020

CITATIONS
0

READS
16,082

2 authors:

Sharnil Pandya
Linnaeus University Sweden
93 PUBLICATIONS **2,951** CITATIONS
SEE PROFILE

Pooja Mehta
Gandhinagar Institute of Technology
7 PUBLICATIONS **247** CITATIONS
SEE PROFILE

All content following this page was uploaded by Pooja Mehta on 06 October 2020.

The user has requested enhancement of the downloaded file.

A Review On Sentiment Analysis Methodologies, Practices And Applications

Pooja Mehta, Dr.Sharnil Pandya

Abstract: The Sentiment Analysis is sometimes a technique to look at the information that is the form of text and determine opinions content from the text. It is also termed as emotion or feeling mining. On-line communication channels like Twitter, Facebook, YouTube, and so forth are these days a lot of passion into human life. People share their thoughts or feelings thereon. During this review paper, we tend to match on opinion mining or feeling assessment which is an area of web data mining and Machine Learning. This paper shows aftereffect of examination by utilizing different ML and Lexicon investigation methodologies. Outcomes are analyzed to play out an evaluation study and check the estimation of the present composition. In this manner, it will help the future investigators with understanding present beginnings in the configuration of possibility examination.

Index Terms: Sentiment Analysis, opinion, emotions, Machine Learning, Accuracy, NLP, support vector machine

1. INTRODUCTION

Human conduct is incredibly affected by their abstract sentiments and convictions, for example, demeanor, feeling, assessment or slant. The choices we make can be affected by others' impressions of the world to an extensive degree, on the grounds that conveying the others' assessments is wired into every individual normally and portrays us as 'social creatures'. Nowadays, social communication channels like Twitter, Facebook, and YouTube have obtained so much popularity. Opinion mining is the other name of Sentiment Analysis which is under the category of machine learning and data mining. From the use of different social media, opinion mining or sentiment analysis techniques have to start with people's data for the analysis of a different kind of area like politic, economy or biology, etc. [2]. Massive amount of information related to distinct individual entities are recorded every day in digital forms. And hence such a fast growth of the field co-exist along with other social media related stuff such a forums discussion ,blogs, customer reviews ,Twitter and social network sites. Sentiment analysis includes classification of data into various classes like optimistic i.e. good sense or negative i.e. bad sense or neutral i.e. non-effective. Sentiment analysis is the task of perceiving whether a given opinion is positive or negative in general[3](e.g., a movie review, a person, a political party, or a policy or product feature review. Because of the free format of messages and easy accessibility of micro-blogging platforms, most of the data on social media are unstructured [3]. When it is necessary to make conclusion or final output, it is important to get Opinions of persons. From different people's experience and reviews which contain important resource. Recent work have been done in promotion and challenging areas with the implementation of opinion mining which serves desires of the individuals [5]. Appropriate training set is required for sentiment analysis for better performance, and accurate dataset for improper analysis of the text. For better means and accuracy, the linguistic analysis is considered Machine learning. The staying of this paper is organized as the followings: The next segment portrays explain about the introduction of the sentiment mining and opinion mining. From that point onwards, different stages of opinion Analysis are displayed. The portrayal of the previous work that is in Section IV was done on sentiment analysis techniques is available. Introduction about the sentiment analysis sources/ resources are in Section V. Part VI explained regarding challenges of sentiment analysis. Finally, the Discussion and conclusion is stated in section VII

- *Pooja Mehta is currently pursuing PhD degree in Faculty of Technology and Engineering in computer engineering in C. U Shah University, Wadhvan ,Surendranagar ,Gujarat,India. E-mail:poojamehta810e@gmail.com*
- *Dr.Sharnil Pandya is currently Associate Professor in Symbiosis University, Pune, Maharshtra ,India, E-mail: sharnil.pandyae@sitpune.edu.in*

A .SENTIMENT:

Sentiments can be expressed in different ways. It can be expressed with various kinds of emotions, by passing judgments, vision or insight or views of people. An emotion can be expressed as sudden reaction by person consciously or unconsciously depending on situation. If we study the emotion in text format then it can be seen in two different ways. First it can be studies as impact on writer – the way he chooses word to express certain emotions. And second impact on reader-how he interpret content written depending on his state of mind and his ability to analysis things.

B Sentiment Analysis:

Sentiment analysis or opinion mining is the process to identifying and recognize or categorized the users' emotion or opinion for any services like movies, product issues, events or any attribute is positive, negative or neutral. The sources for this analysis is social communication channels i.e Web site which include reviews, forum discussions, blogs, micro-blogs, Twitter etc. This research field is very popular nowadays because of its opinioned data where user can find reviews for any services which are useful for their daily lives. The large amount of opinionated data is stored in digital forms. For particular topic or any opinion the sentiment analysis which relates the mining of the data works and give the output. For sentiment analysis, research works on emotion-based summarization, feeling or mind extraction. Sentiment analysis also known as Opinion mining which uses NLP – Natural Language Processing to following the emotions, feelings of the public opinion about a particular topic for any product or services. As Sentiment analysis is very famous, it can be also useful in many ways in surveys and advertisement campaign by getting the success rate of any product or services with people's opinion or suggestion. .It also give the information about people liking and disliking and company gets much clear idea regarding its product features. Sentiment Analysis has

increased a lot of acceptance among various zone like politics [9], business [10] and marketing/selling and advertisement (to estimate sales of specific products). So identifying type of sentence is the most important part of opinion mining. We have to classify the sentence either subjective or objective. Recent or existing research is using both supervised and unsupervised learning technique to provide different techniques for several purpose of sentiment analysis. In initial research all or combination of below supervised techniques are used.

1. Support vector machine
2. Maximum Entropy
3. Naive Bayes

Unsupervised techniques used by intial research are

1. Exploit sentiment lexicons
2. Grammatical analysis
3. syntactic patterns

In general, Sentiment Analysis includes advanced processes. The analysis has a totally different series of tasks, sentiment classification (supervised or unsupervised), subjective or objective analysis, and opinion extraction. For any text document or a sentence, to do the labeling as subjective or objective can be evaluating by subject level analysis. The sentiment classification includes the duty of probing the sentiment polarity of the filtered sentences . All the sentences are dividing in to neutral, negative or positive type which is depends on the emotions we get from the reviews.

1. Subjective / Objective Classification

One of the main tasks in sentiment analysis (SA) are Subjective and objective classification. The main purpose is to divide the entire documents or sentences into one of the two classes; objective or subjective [11]. In Opinion mining, the sentences in the text are labeled as either subjective or objective. By using the sentiment analysis we can extract subjective sentence. The factual information generated by objective sentences need to be removed. When it contains emotions or feelings subjective sentences are important to the procedure with sentiment analysis. Subjective sentence are made up with following views of users, perspectives, thoughts, comments and opinions about the sentence level.

Example-

1) Subjective: Inception is an awesome movie.
(The sentence having emotion (awesome), thus it is subjective)
2.) Objective: Leonardo is the actor of Inception. (there is no sentiment in this sentence, it is an actual fact, and thus it is objective)[13].
For subjective sentence various research are going on.For any organization ,objective lines will be eliminated and subjective sentence will be consider in research as it contain the opinion and emotion. English language is mostly uses for the subjective classification research but other various studies being done in many languages .Arabic language and urdu language are one of them with the use of different supervised learning classification methods and support vector machine (SVM) is one of them for the subjectivity and sentiment analysis.

C CLASSIFICATION LEVEL

There are different types of classification levels: 1) Document-level 2) the Sentence level 3) Attribute /Aspect Level

1) Document Level: This is very first level of Opinion mining or sentiment analysis which is only based on the document. In this particular level, we take the whole document is taken into consideration and figure out the polarity. Through this level or with the help of this level we can classifies whether the available opinion or emotions provide us a positive sentiment or negative sentiment [4].To consider this, the document should be on a single topic. The main source of this document to be considered is sentiment or emotions. For example, in one text, the file contains the review of only single product, now that the system starts calculating whether the whole review is expressing an overall positive or negative opinion about this product. Thus, for many products review this type of level is not valid. The main and only advantage of is that we get most of all polarity of a particular feature and the drawback of this level is that people's liking and disliking didn't get by this.

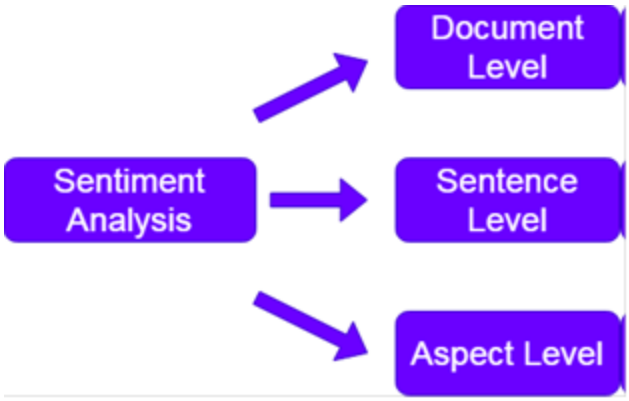

Fig.1 *Categories of Sentiment Analysis*

2) Sentence Level: One of the other categories of sentiment analysis is sentence-level, in which to determine the polarity of each sentence is processed and analyzed which gives a positive, negative or neutral opinion about the sentence. Subjective sentence are made up of views of users, perspective and opinion about the sentence. Sentence is neutral when it does not imply any opinion. When sentence is neutral it is more likely to be called objective sentence which gives fact information under subjective classification called subjective sentences which shows subjective views and opinions. Subjectivity and objectivity classification are the main advantage of sentence analysis. Generally, subjective sentences is detected in machine learning. But, at the sentence level, there is a limitation with sentiment analysis

3) Aspect/Feature Level: The other name of the Aspect level is also Feature level or entity level. We can find people's likes and dislikes in both the document and the sentence level analysis. Finally the output of this Aspect level which express the output as opinion is most detailed type of sentiment analysis. Two outcomes are considered POSITIVE or NEGATIVE and the target value. To find out sentiment on entities and their features, target opinion helps to realize the importance of this level. Reviews, feedbacks, comments, and complaints, etc. are performed at this level.

2. APPLICATIONS

A. Decision making support:
Building a website that could perform decision making is a very crucial part. Analysis has its own advantage like; it can lead to different ideas which can help us to make decision in day to day life such as choosing a good restaurant to go for dinner, or buying a new car or selecting a good movie to watch etc.

B. Business related application:
Because if every day changing market, the competition has increased a lot in co-operative world. Every wants to create a innovative and newest product which can fully satisfy their customers. To achieve more valuation of their product, organization can assemble all the needs of their users and enhance the efficiency of product from feedback collected from their customers.

C. Predictions and trend analysis:
Tracking views of public by sentiment scrutiny which enable any person to predict the market scenario which helps any person for trading and polls market. By using this all opinions user can predict the market trends.

3. MATERIAL AND METHOD AND APPROACHES

Numerous methodologies are available for opinion mining, but two main groups are used. The problems of SA will be solved by the first group using by implementing the machine learning approach. The second group uses lexicon-based method which is a linguistically-inclined method. In both groups, many techniques exist. From the following way, we can extract the features of text or sentences.

1) N-Gram: Only one word can be taken by one at a time (unigram) or two words (bigram) up to n words as a result. Unigram features cannot be captured by some opinions. For example, this book is fascinating. It is an optimistic comment if in only unigram model it is fascinating to take it together and negative.

2) POS tagging: –"It is the way of words to signify it in content (corpus) as it is linked to its parts of speech in the light of both its definition and its connotation with touching the words. Nouns, pronouns, adjectives, adverbs, etc. are examples of different parts of speech".

3) Stemming – In this, eliminating prefixes and suffixes is the main process.. For example, 'running', 'sleeping', 'ran' can be stemmed from 'run' and 'sleep' respectively. It basically helps in Cataloging but sometimes it also leads to decrease in cataloging accuracy.

4) Stop words – Stop words are Pronouns (he/she, it), articles (a, an, the), prepositions (above, in, near, under, besides). These words are nothing but offer no or little information about the emotions. On the internet you can access list of stop words. In the pre-processing step, it can be used to remove them.

5) Conjunction handling - In general, there is only one meaning of each sentence at a time. But there are certain available conjunction words like But, And, while, although, however, changes the whole denotation of its sentence. For example, even though the ride was good but it was not up to my hopes. By using these rules throughput can be amplified by 5% [6].

6) Negation handling - Negation words like 'not' inverts the gist of the whole sentence. For example, the movie was not good as 'good' in it which is optimistic but 'not' upturns the schism to negative.

To identify emotions or opinion words is an important task in many applications in opinion mining. From the given feature ,classifying the polarity is basic important task. Positive, Negative and Neutral are three classes where the polarity is categorized .From Polarity identification, calculation of sentiment strength, sentiment score etc. can be done using Lexicon techniques. There are various ways and techniques are available for opinion mining, there are majorly two groups used. 1) Uses lexicon methods and 2) machine learning method which resolves the problems of SA.

1)Lexicon based approach: In this current approch ,when using the available lexicon techniques for a text which is given, will separate the words.In general it performed by aggregation of scores : for example subjective words scores as positive,negative and nuetral etc are summed up separately for same. It assigns a score to each word ..Atlast four scores are generated . The one which gets the maximum score gives the overall split of the text[10]. It has mainly divided into two parts.a)Dictionary-based b)corpus-based.

A) Dictionary-based approach - In this system,the user collected a set of sentiments words and seed list is prepared by them.After that ,the user start searching for phrasebooks and lexicon to find synonyms and antonyms of particular text.Once this is done,the newly created substitutes are added into the seed list.Untill there are no new words are found to users this process continues..

Disadvantage: There has to have struggle in finding context or domain-oriented emotion words.

B) Corpus-based approach - Corpus is a basically a term which is a cluster of writing like group of some writing which is often on a very precise matter.In this,users uses the help of corpus text to drawn-out the seed list which is in organized situation[9].

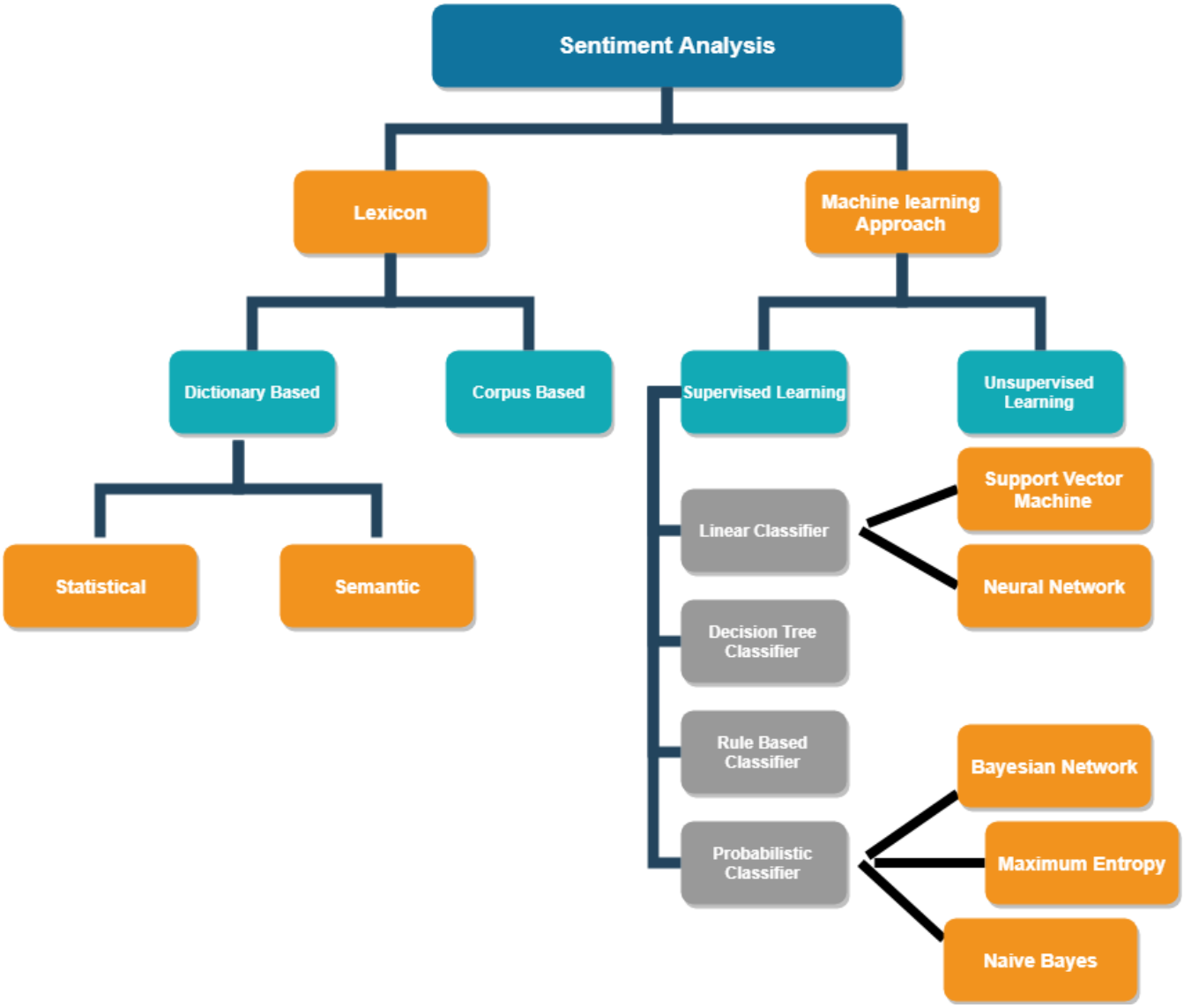

Fig. 2 *Various Approaches of Sentiment Analysis*

2) Machine Learning approach:

In this approach, initially classification is performed by taking two different assemblies of the document. Trained data and test data are part of these. This is termed as involuntary classification. Further text is extracted from the features and categorized into I) supervised and II) unsupervised.

a) Supervised System- Among various kind of datasets ,Labeled training dataset is one of them which is used in supervised system. Each type of class has its own property and advatanges and has its label related to it which can be used for this system.Each word ,upon arriving is categorized under a label depending on its type and charecteristics related with it.

1. Probabilistic classifier: Predicts or anticipates probability function related to input records among different modules.

a) Naïve Bayes: In this, to generate possibilities of a group to provide prediction that group of properties belongs to one particular label with help of Bayes theorem using merely a text document as an input. BOW - Bag of Words is a way to extract a text with using machine learning methologies which is simple and easy to implement. This existing model conduct that these all the features which are given autonomous. [20]

P (label/features) =P (label)* P (features/label)/P (features)

b) Bayesian Network: It is used to manifest relationships among different features.It can be compared with acyclic graph in which nodes represent random variable and edges represent dependencies This model is very pricey and hence it's hardly used.

c) Maximum Entropy: By doing encoding, the labeled feature sets are converted into to vectors by using classifiers .This vector are converted and utilized to decide the weights of these features which can be able to use to suppose and predict the label for each of their feature set.

2. Linear classifier: The characteristics of the linear classification can be implementing by using this classifier which is used to shows predictor as result and can be divided in to two classes:

A) Support Vector Machine (SVM): This learning model is under supervision to utilize for classification. The most important purpose of this particular model is to assure that this is the best linear separator for classification. This will make a model that results in new information into one or two classes using SVM training.

B) Neural Network (NN): It is a neural structure of the brain having electronic networks of neurons. In this network, Neuron

is the basic component. Neurons are categorized in to three parts- input, hidden and output.

3 .Decision tree classifier: To make division of the data, there is a condition which is used. one class consist those data which mollify the condition and other class consist of the remaining of the data. This technique is called a recursive technique which has two parts: single attribute split and multi attribute split.
4. Rule based classifier –It is condition based classifier which makes usage of condition or rule like IF, THEN. It can be written as
IF condition THEN decision
We can produce the rules based on our requirements at the time of training phase [2].

4. RESOURCES OF SENTIMENT ANALYSIS

To collect data is the main purpose of Sentiment analysis where social communication channels like Twitter, Facebook or any pre-existing resources.
A) Blogs & Forums: It is source of opinions and emotions where we get information for research purpose and that all information can be used by researchers via Web forums and blogs. Generally, for only single subject forums are designed; thus, by using the forums we can ensure the sentiment mining in single domain. As well, it's the trend that bloggers updated their blogs and reviews every day after activities in and around their areas, countries and around the world.
B) Reviews: There are many available studies which dedicated only on reviews because of their usability with the opinions and sentiment. During any research, Movie and product reviews were mostly studied by researcher where the main purpose is to get the feedback from the sentiment and opinions.
C) News Articles: News articles, such as financial articles and political reviews are a popular source of sentiment analysis [51] . The main format of News articles texts is structured and formal.
D) Social Networks: Many social networks sites are available from which we can take the opinions and reviews for sentiment analysis like Twitter, Facebook, etc.

- Twitter:

Tweets are the messages posted by different users, having restriction of 140 characters. Users can read message (called Tweets) of one another. The micro-blogging service which provides this facility is knows as Twitter. By using this tweets which can work as opinions and reviews for future patterns where we can generate the poll results.

- Facebook:

The provision of posting personal profile, photos, videos and other related information are provided by most famous social networking facility called Facebook which is popular right after it got launched in 2004.

Hence, these much ample amount of information available in form of user's message, computer technology which is dependent on sentiment behind this message is introduced known as sentiment analysis.

5. RELATED WORK

A lot of studies have been done by scholars to analyze emotions or opinions. There are many methods used to extract the data.

Comparison of various Approaches and Methods

In [22] and [24] research work used a feed forward neural network for:
1) Identifying online users who express their feelings, show perspective and tweets oftenly. And 2) To characterize these tweets in different categories based on positive and negative keywords, For this purpose they also used Twitter API. Convolutional Neural network is one of the methods which are used for Sentiment Analysis. By combining sentiment analysis and the Morphological Sentence Pattern Model we can get many good outcomes. The other techniques like Tokenization, stemming, and preprocessing, self- organizing map (SOM), and a recursive neural network can also be used for sentiment analysis. In [21], the proposed system shows that after data preprocess it classifies into sentence-level and then extracts the features of the data. After then it applies the coherence resolution and usage of SentiWordNet. .It applies the SVM machine learning approach to count the accuracy of the product feature. At last, we find the overall sentiment or accuracy regarding the feature of the product. In [23], research works shows that different machine learning methods are used to extract the emotions. It uses the different twitter data to preprocessing, subjective classification and feature classification. At last, it counts the accuracy for all machine learning methods like SVM, Naïve Bayes and ME. In [25], proposed different method which works on artificial intelligence .Propose model works on VADER method which is different than tradition methods like SVM or ME. By using VADER method which is Valence Aware Dictionary and sentiment Reasoner the opinions are categorized in to positive, negative and neutral The user emotions are divided in to optimistic, negative and unbiased by a. The result shows the highest rank of three artificial intelligence. The best resource for the SA is review data. In [26] researchers uses techniques of NLP and computational linguistics to classify the sentiments of the reviews of the hotel data. This outcome shows the result of satisfaction, security, comfort, luxury and lodging services for tourist person. It would help the hotel managers to have what customer needs, discover areas for further development and increase its service quality. NLP techniques are used for reviews data which works with the processing of textual data for sentiment analysis. Here, researchers used Sentimental polarity based model (SPBM) for their work. It uses multinomial algorithm from Naïve Bayes method which gave good prediction results when compared to other classification algorithms. In Opinion mining, E-commerce and news type datasets are available. In [27] and [28] research, they took the data set from Amazon and BBC – online news channels where they proposed works. While handling in all the datasets, it divides positive and negative text for the objective, from the features review and articles respectively and performed different analysis methods like preprocessing where data is cleaned for analysis. At first, the goal is to calculate the polarity of textual data whether it is positive or negative. Naïve Bayes and SVM methods are used to find the accuracy and precision of the data which are supervised learning method. Collection of known and defined words are called Sentiment lexicon. Two types of sentiment analysis are :1) Lexicon Based & 2) Machine learning .Polarity shift is the main concern in any aspect or feature level. And research such as [29][31], are done to find this polarity shifts .In their research Bag-of words which handle text data as vector of different words. And various ML techniques are used to categorize these words.

This model also lacks majorly with polarity shift issue .Technique called PSDEE-Polarity Shift Detection, Elimination, and Ensemble is used to address this polarity shift issue in Document level sentiment analysis. And then ML techniques such as Naïve Bayes and SVM are used for sentiment polarity (which is positive, negative or neural) after preprocessing the data. Twitter site is a micro-blogging because it's data are not in structured format. These data are shared by different users in form of their feelings or about their daily life. Twitter data are appropriate in data stream mining as data or messages are small and continuous. [30][32][33][36][41], research works on twitter data which analysis sentiment from its short text. Text data can be categorized into positive or negative opinion .In general or specific item feature is called sentiment classification. Views of different people about specific product can be divided into positive sentiments, negative sentiments or neutral sentiments. Supervised machine learning method is most reliable method for sentiment analysis. The machine learning algorithms which are useful for sentiment analysis are Naive Bayes, Maximum Entropy and Support Vector Machine (SVM) to find the accuracy of the product feature. Sentiment Analysis is a very challenging and important task that works on machine learning . In [35], propose sentiment classification works on the Arabic language. They spoke to that Arabic tweets represent a decent open door for opinion mining research however they were postponed because of lack of sentiment analysis assets or difficulties in Arabic language text analysis. There are two levels available for classification in Arabic. In first one subjectivity analyzer based on supervised approaches and filters the reviews in relevant and irrelevant and second level sentiment analyzer based also on supervised approaches and ensemble techniques to classify relevant reviews into positive, negative and neutral. For utilizing diverse weight plans, stemming and n-grams procedures tests were led which demonstrated that SVM classifier utilizing TF-IDF through bigrams includes was better when contrasted with Naive Bayesian classifier. Naïve Bayes (NB), Support Vector Machines (SVM) and Rocchio classifiers including in this classification. In [34],they proposed sentiment analysis for dealing with any topic from related documents which gives the output as positive or negative. For opinion retrieve, the topic-related structured are made with the help with query-dependent. To implement ranking algorithm for data retrieval researchers uses SVMRank. To calculate performance of ranking, many methods that can be used to. Mean Average Precision (MAP) as the evaluation metric used for TREC community. In [37], by using the Naïve Bayes classifier it can detect the polarity of the English tweets whether tweets are positive, negative or nuetral. Two unique variations of Naive Bayes classifiers were constructed 1)specific Baseline and 2) Binary (which makes use of an lexicon methods and groups as positive and negative) Multiword from various sources and Valence Shifters are identified by this approach. In [38] talks about the social media site like twitter, Facebook which is very famous in social media networks. They propose a new framework to finding the polarity of the opinion or emotions from the web dataset. It joins this system with manually data from Twitter. Twitter API is used to gather data. It analyze the data in to positive , negative and neutral. Unigram Naive Bayes which is sub method of Naiva bayes approach is used for this. In [39], they propose a supervised sentiment classification framework which is based on data from Twitter to find the accuracy of the data. For Twitter client characterized it include hash tags in tweets , single words, n-grams which are then consolidated into a solitary element vector for sentiment order. K-Nearest Neighbor algorithm is used to allocate sentiments names by building an element vector for every model in the preparation and test set.

Table 1.0 *COMPARATIVE STUDY OF TECHNIQUES OF SENTIMENT ANALYSIS*

Sr.No	Year	Paper Title	Methodology Used	Review Dataset	Accuracy
1	(2017)	A Feature Based Approach for Sentiment Analysis using SVM and co-reference Resolution[21]	SVM & co-reference Resolution	Training Dataset of Product Review	73.6%
2	(2015)	Neural Networks for Sentiment Analysis on Twitter[22]	Neural network with Feed Forward Method	Twitter Dataset	74.15%
3	(2017)	Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python[23]	Naïve Bayes, SVM Maximum Entropy	Twitter Dataset	86.4% 73.5% 88.97%
4	(2018)	Sentiment Analysis using Neural Networks: A New Approach[24]	Convolutional Neural Network	Product Data Review Twitter Data	74.15% 64.69%
5	(2018)	Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants[25]	Valence Aware Dictionary and Sentiment Reasoner (VADER)	Reviews of Electronic product	87.4%
6	(2018)	A framework for sentiment analysis with opinion mining of hotel reviews[26]	Naive Bayes	Hotel Reviews from OpinRank	83.5%
7	(2018)	Aspect-Level Sentiment Analysis on E-Commerce Data[27]	Naïve Bayes SVM	Amazon Customer Review Data	90.423% 83.43%
8	(2017)	Document Level Sentiment Analysis from News	Machine learning approaches	BBC News Dataset	57.7%

		Articles[28]			
9	(2017)	Polarity Shift Detection Approaches in Sentiment Analysis: A survey[29]	Lexicon-based and Supervised Machine Learning-based	Product Review	84.6%
10	(2017)	A Sentiment Analysis Method of Short Texts in Microblog[30]	Language Technology Platform (LTP) for dependency syntax analysis	COAE2014(BBC DataSet)	86.5%
11	(2016)	SemEval-2016 Task 4: Sentiment Analysis in Twitter[32]	SVM	Twitter Dataset	84.5%
12	(2016)	A Topic-based Approach for Sentiment Analysis on Twitter Data[33]	SVM	Twitter Dataset	74.09%
13	2013	Ensemble of Classification Algorithms for Subjectivity and Sentiment Analysis of Arabic Customers' Reviews[35]	Naive Bayes, SVM	Arabic Reviews from jeeran.com(service and product reviews)	97.06% 89.1%
14	(2014)	Cities: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets[37]	Naïve Bayes	Training Dataset of Tweets by SEMEVAL2014	76.54%
15	(2013)	Opinion Mining on Social Media Data[38]	Naïve Bayes	Twitter Dataset	76.8%
16	(2010)	Sentiment Knowledge Discovery in Twitter streaming Data[41]	Multinomial Naïve Bayes	Twitter API	82.45%
17	(2010)	Twitter as a Corpus for Sentiment Analysis and Opinion Mining[42]	SRF	Twitter Dataset	56.4%

In [40], they proposed to method for 3-route show for ordering sentiment into positive, negative and neutral classes. Unigram based model and a tree kernel based model are used tor this analysis. For tree kernel-based model ,it made with spoken tweets which works as tweets - tree. 100 highlights uses by highlight based model and unigram uses more than 10,000 highlights. The tree kernel-based model beat the other two models. In reference[41], Firehouse API , which gives Twitter streaming data to fetch all messages from every users which is openly available. Different methods such as stochastic angle plunge, the Hoeffding tree and multinomial naive Bayes are used to test these data. They concluded that when SGD based model is used with appropriate learning rate ,it is better than other methods. Micro-blogging websites are rich sources of data for opinion mining and sentiment analysis. In [42], A model has been suggested to categorize message of user (aka tweets) in positive and negative. We recommend ways to gather mass of data which can be used in sentiment analysis and opinion mining purposes. In this ,they have used Twitter API to collect massive amount of tweets which uses emotions and subsequently commenting on them. Relied on the techniques such as multinomial naïve Bayes which uses N-grams and POS labels and using this massive amount of tweets .They created a sentiment classifier. Since dataset was used just having tweets with emoticons .It was not quite up to the level effective. A successful sentiment classification based on the rich data set which could provide unique data for business, politics or product feature, etc. In [43], to distinguish tweets they used two models a naïve bayes Bayes classifier's performance was far better than Maximum Entropy demonstrate. In [44], A distant supervision technique which is having dataset made up of tweets with emoticons, was used for sentiment analysis of twitter. They used model utilizing naïve bayes , MaxEnt and support vector machine (svm) altogether whole element space comprised of unigram ,bigram and POS .In all of these SVM was superior among all.

In [45], propose a method for tweets ordering by using a two-stage programmed sentiment analysis In first phase they collect tweets and arrange it as objective and then afterward in the second phase , the abstract tweets were named positive or negative. The element space utilized included re-tweets, hash-tags, link, accentuation and exclamation stamps related to highlights like the earlier extremity of words and POS.

6 CONCLUSION & FUTURE SCOPE

In this paper, we mainly focus on the basics of sentiment /opinion mining and its levels. There are various approaches and methods to identify sentiment from content. In this paper, our examination represents machine learning procedures. From various classification methods, Sentiment Analysis indicates the results into positive, negative and neutral scores. The study shows that machine learning methods, such as SVM, Naive Bayes, and neural networks have the highest accuracy and can be considered as the baseline learning methods as well as in some cases lexicon-based methods are very effective. In future work, discovering the result of various other combinations of text data and other on prediction accuracy can be done. More work in the future is needed to improve performance measures.

REFERENCES

[1] Surnar, Avinash, and Sunil Sonawane. "Review for Twitter Sentiment Analysis Using Various Methods."IJARCET-VOL 6-ISSUE 5,2017.

[2] Eliacik, Alpaslan Burak, and Erdoğan Erdoğan. "User-weighted sentiment analysis for financial community on Twitter." Innovations in Information Technology (IIT), 2015 11th International Conference on. IEEE, 2015.

[3] Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani|, Veselin Stoyanov."SemEval-2016 Task 4:Sentiment Analysis in Twitter", Proceedings of SemEval-2016, Association for Computational Linguistics.

[4] Ahmed, Khaled, Neamat El Tazi, and Ahmad Hany Hossny."Sentiment Analysis over Social Networks: An

Overview." Systems, Man, and Cybernetics (SMC), IEEE International Conference on. IEEE, 2015.
[5] Rasika Wagh,Payal Punde." Survey on Sentiment Analysis using Twitter Dataset" Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Xplore ISBN:978-1-5386-0965-1
[6] Anchal Kathuria, Dr. Saurav Upadhyay." A Novel Review of Various
Sentimental Analysis Techniques" International Journal of Computer Science and Mobile Computing, Vol.6 Issue.4, April- 2017, pg. 17-22.
[7] D. M. E.-D. M. Hussein, "A survey on sentiment analysis challenges," J. King Saud Univ. - Eng. Sci., vol. 34, no. 4, 2016.
[8] Liu, B. Sentiment analysis: mining opinions, sentiments, and emotions. The Cambridge University Press.2015.
[9] Bilal Saberi, Saidah Saad."Sentiment Analysis Or Opinion Mining: A Review".International Journal of Advanced Science Engineering Information Technology, Vol-7(2017), ISSN:2088-5334.
[10] J. Bollen, H. Mao, and X. Zeng "Twitter mood predicts the stock market". Journal of Computational Science, 2(1): 1-8 2011.
[11] T. Xu, Q. Peng and Y. Cheng. "Identifying the semantic orientation of terms using S-HAL for sentiment analysis". Knowledge-Based Systems, 35: 279-289, 2012
[12] T.T. Dang, N. T. X. Huong, A.C. Le and V.N. Huynh."Automatically Learning Patterns in Subjectivity Classification for Vietnamese". Knowledge and Systems Engineering. Springer, pp. 629-640, 2015.
[13] Arora, Piyush. "Sentiment Analysis for Hindi Language." Diss.International Institute of Information Technology Hyderabad, 2013.
[14] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff and S. Patwardhan. "Opinion Finder: A system for subjectivity analysis". In Proceedings of hlt/emnlp on interactive demonstrations, pp. 34-35.
[15] E. Riloff, J. Wiebe and W. Phillips. "Exploiting subjectivity classification to improve information extraction". In Proceedings of the National Conference On Artificial Intelligence, pp. 1106.
[16] P. D. Turney. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews". In Proceedings of the 40th annual meeting on association for computational linguistics,pp. 417-424.
[17] Math Alrefai, Hossam Faris, Ibrahim Aljarah ." Sentiment analysis for Arabic language: A brief survey of approaches and techniques".2018
[18] Emma Haddia, Xiaohui Liua, Yong Shib, ''The Role of Text Preprocessing in Sentiment Analysis'', ELSEVIER, Procedia Computer Science 17 (2013) 26 - 32.
[19] Jagdale, Rajkumar S., Vishal S. Shirsat, and Sachin N. Deshmukh. "Sentiment Analysis of Events from Twitter Using Open Source Tool." (2016).
[20] Kang Hanhoon, Yoo Seong Joon, Han Dongil., "Senti-lexicon and improved Naï¨ve Bayes algorithms for sentiment analysis of restaurant reviews", Expert Syst Appl ,39:6000–10, 2012
[21] Hari Krishna M, Rahamathulla K, Ali Akbar," A Feature Based Approach for Sentiment Analysis using SVM and Coreference Resolution", International Conference on Inventive Communication and Computational Technologies,ICICCT2017.
[22] Brett Duncan and Yanqing Zhang, "Neural Networks for Sentiment Analysis on Twitter", IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCICC 2015)
[23] Monika Negi, Kanika Vishwakarma, Goldi Rawat, Priyanka Badhani, Bhumika Gupta," Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python", International Journal of Computer Applications (0975 – 8887) Volume 165 – No.9, May 2017
[24] Shiv Dhar,S.Pednekar,K.Borad,Prof.Ashwini Save," Sentiment Analysis using Neural Networks: A New Approach", International Conference on Inventive Communication and Computational Technologies (ICICCT 2018)
[25] Chae Won Park, Dae Ryong Seo, "Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants", 5th International Conference on Industrial Engineering and Applications,2018.
[26] Kudakwashe Zvarevashe, Oludayo O. Olugbara," A framework for sentiment analysis with opinion mining of hotel reviews", Conference on Information Communications Technology and Society (ICTAS) 2018.
[27] Satuluri Vanaja, Meena Belwal," Aspect-Level Sentiment Analysis on E-Commerce Data", International Conference on Inventive Research in Computing Applications (ICIRCA 2018).
[28] Vishal S. Shirsat, Rajkumar S. Jagdale, S. N. Deshmukh,"Document Level Sentiment Analysis from News Articles", International Conference on Computing, Communication, Control and Automation (ICCUBEA) 2017.
[29] Sayali Zirpe, Bela Joglekar," Polarity Shift Detection Approaches in Sentiment Analysis: A survey", International Conference on Inventive Systems and Control,2017.
[30] Jie Li; Lirong Qiu," A Sentiment Analysis Method of Short Texts in Microblog", International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC) 2017.
[31] Erik Cambria, Nanyang Technological University,"Affective Computing and Sentiment Analysis", IEEE Intelligent Systems,2016.
[32] Preslav Nakov|, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani|, Veselin Stoyanov," SemEval-2016 Task 4: Sentiment Analysis in Twitter", Proceedings of SemEval-2016.
[33] Pierre FICAMOS, Yan LIU," A Topic based Approach for Sentiment Analysis on Twitter Data", International Journal of Advanced Computer Science and Applications 2016.
[34] Zhunchen Luo , Miles Osborne ,TingWang," An effective approach to tweets opinion retrieval", Springer Science+Business Media New York 2013.
[35] Nazlia Omar,Mohammed Albared, Adel Qasem Al-Shabi,Tareq Al-Moslmi," Ensemble of Classification Algorithms for Subjectivity and Sentiment Analysis of Arabic Customers' Reviews", International Journal of Advancements in Computing Technology(IJACT),2013.
[36] Neha Upadhyay, Prof. Angad Singh," Sentiment Analysis on Twitter by using Machine Learning Technique" ,

International Journal for Research in Applied Science & Engineering Technology (IJRASET),2016.
[37] Pablo Gamallo, Marcos Garcia," Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets", International Workshop on Semantic Evaluation (SemEval 2014).
[38] Po-Wei Liang, Bi-Ru Dai," Opinion Mining on Social Media Data", IEEE 14th International Conference on Mobile Data Management ,2013.
[39] Dmitry Davidov, Oren Tsur, Ari Rappoport," Enhanced Sentiment Learning Using Twitter Hashtags and Smileys", Coling 2010, Beijing, August 2010.
[40] Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau," Sentiment Analysis of Twitter Data".
[41] Albert Bifet and Eibe Frank," Sentiment Knowledge Discovery in Twitter Streaming Data", International Conference on Discovery Science 2010.
[42] Alexander Pak, Patrick Paroubek," Twitter as a Corpus for Sentiment Analysis and Opinion Mining", International Conference on Language Resources and Evaluation, LREC 2010.
[43] Ravi Parikh and Matin Movassate," Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques",2009.
[44] Alec Go, Richa Bhayani, Lei Huang," Twitter Sentiment Classification using Distant Supervision"2009.
[45] Luciano Barbosa, Junlan Feng," Robust Sentiment Detection on Twitter from Biased and Noisy Data", Coling 2010, Beijing, August 2010.
[46] Q. Ye, Z. Zhang and R. Law. "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches". Expert Systems with Applications, 36(3): 6527-6535,2009.
[47] A. B. Goldberg and X. Zhu," Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization". In Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing, pp. 45-52.
[48] R. Prabowo and M. Thelwall." Sentiment analysis: A combined approach". Journal of Informetrics , 3(2): 143-157, 2009.
[49] J. Liu and S. Seneff." Review sentiment scoring via a parse-andparaphrase paradigm". In Proceedings of the Conference on Empirical Methods in Natural Language Processing: Volume 1- Volume 1, pp. 161-169, 2009.
[50] Y. Chen and J. Xie "Online consumer review: Word-of-mouth as a new element of marketing communication mix". Management Science, 54(3): 477-491, 2008.
[51] M. Abdul-Mageed, M. T. Diab and M. Korayem "Subjectivity and sentiment analysis of modern standard Arabic", In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2,2011.
[52] S. Yu and S. Kak. "A survey of prediction using social media". arXiv preprint arXiv:1203.1647, 2012.
[53] L. Pan. "Sentiment Analysis in Chinese". Brandeis University, 2012.

View publication stats